Designation: ~~E1958 – 21~~ E1958 – 22

Standard Guide for
Sensory Claim Substantiation[1]

This standard is issued under the fixed designation E1958; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

INTRODUCTION

Formats or standards for testing related to sensory claim substantiation cannot be considered without a frame of reference of where that format or standard would fit within the legal framework that surrounds the topic. Product sensory claims tests are performed for three basic reasons: (*1*) *Comparison of Products*—Determines how one product compares to another, usually a competitor or earlier version of itself. (*2*) *Substantiation of Claims*—Enables marketing personnel to use positive references through advertising or packaging, or both, in the presentation of the product to the consumer. (*3*) *Test Performance*—Ascertains and establishes the tested product performance within the scope of its intended use.

The risk associated with each claim is assessed when considering claims substantiation. Compelling and aggressive claims are sure to be scrutinized closely by competitive firms, and if inconsistencies are found through competitive test data, the claims could be challenged in one or more of the following venues: *(1)* National Advertising Division (NAD) of the Advertising Self-Regulatory Council (ASRC), *(2)* one or more media, such as print, broadcast, or electronic media, *(3)* Consumer Advocacy Organizations, and *(4)* Civil or Federal courts. No single test design or standard test will prevent challenges. The criteria used by each of the potential forums are not identical and are constantly evolving. With the introduction of new technologies coupled with changing consumer demands, testing processes and protocols that were sufficient five or ten years ago may not hold up under today's criteria and scrutiny. Testing requirements of the future can only be a matter for speculation. The one constant is that, as advocates of their clients' positions, attorneys will defend their clients' testing processes and protocol while questioning with great detail every aspect of their competitor's protocol in the attempt to sway the arbiter to agree that their clients are in the right. Legal counsel should be part of any team developing claim substantiation.

This guide demonstrates what a group of professionals who are skilled in the science of testing consider appropriate from a scientific and technical standpoint, and represents an effective method for both defendant and challenger to determine the viability of a sensory claim. The key word is "appropriate." If a particular aspect of a test, or method, is not appropriate for a specific application, it should not be used. Care should be taken to clearly define the reasons and data supporting a deviation from the standard, as any departure invites scrutiny. Since departures are inevitable, the word "should" is used in this guide to indicate when other techniques may have applications in certain unusual circumstances. Whenever a test protocol has been completed, it should be critiqued for weaknesses, including whether experts in the relevant field would consider the research objectively designed, conducted, and analyzed, using procedures that give accurate and reliable results. If weaknesses are found, corrective action should be taken, since the competition may point out any weakness or discrepancy and challenge the study.

While the scientific and technical community identifies the appropriateness of a research method

---

used to support a sensory claim, the legal community evaluates substantiation for legal claims using "reasonableness" as the criterion. With the importance of having a legal "reasonable basis" for a claim, the question remains, "What is reasonable?" Unfortunately, there is no specific answer to that legal question, as it will depend on the type of claim, product application and use, applicable regulations where the product is sold, and other factors. These considerations, market pressures (such as timing), and testing budgets can influence and impact the protocols to support a specific claim. This guide provides principles and considerations that need to be addressed for good sensory and consumer testing practices.

1. Scope

1.1 This guide covers reasonable practices for designing and implementing sensory tests that validate claims pertaining only to the sensory or perceptual attributes, or both, of a product. This guide was developed for use in the United States and must be adapted to the laws and regulations for advertisement claim substantiation for any other country. A claim is a statement about a product that highlights its advantages, sensory or perceptual attributes, or product changes or differences compared to other products in order to enhance its marketability. Attribute, performance, and hedonic claims, both comparative and non-comparative, are covered. This guide includes broad principles covering selecting and recruiting representative consumer samples, selecting and preparing products, constructing product rating forms, test execution, and statistical handling of data. The objective of this guide is to disseminate good sensory and consumer testing practices. Validation of claims should be made more defendable if the essence of this guide is followed.

Table of Contents

1.2 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]
  E253 Terminology Relating to Sensory Evaluation of Materials and Products
  E1885 Test Method for Sensory Analysis—Triangle Test
  E2164 Test Method for Directional Difference Test

2.2 *ASTM Publications:*[3]
  MNL 13 Descriptive Analysis Testing for Sensory Evaluation
  MNL 26 Sensory Testing Methods: Second Edition
  STP913 Physical Requirement Guidelines for Sensory Evaluation Laboratories

3. Terminology

3.1 *Definitions:*

3.1.1 Terms used in this guide are in accordance with Terminology E253. Additional terms are listed below.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *α (alpha) risk, n*—the probability of concluding that a sensory claim is supported when, in reality, the sensory claim is false.
  3.2.1.1 *Discussion*—
α risk also is known as Type I Error or significance level.

3.2.2 *β (beta) risk, n*—the probability of concluding that a sensory claim is not supported when, in reality, the sensory claim is true.
  3.2.2.1 *Discussion*—
β risk also is known as Type II Error.

3.2.3 *central location ~~testing~~test (CLT), n*—a test where stimuli are evaluated by consumers at a common designated ~~location(s). The~~location(s); the stimuli preparation and presentation are usually controlled.
  3.2.3.1 *Discussion*—
Examples of designated locations include market research facilities, academic laboratories, grocery stores, or hotel conference rooms.

  ~~3.2.1.1 *Discussion*—~~
~~Examples of designated locations include market research facilities, academic laboratories, grocery stores, or hotel conference rooms.~~

3.2.4 *comparative ~~claims,~~claim, n*—designed to compare similarities and differences between two or more products. The basis for comparison can be within the same brand, between two brands, or between a brand and other products in the category.

3.2.5 *context effect, n*—effect upon the perception of a stimulus arising from its interrelationship with other stimuli, such as in a presentation set or in the experiment.

3.2.6 *directional difference test, n*—a paired comparison or 2-AFC (Alternative Forced Choice) method in which assessors select the stimulus from a pair of stimuli that is perceived to be higher or lower in intensity of a specified sensory attribute.

---

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.
[3] Available from ASTM International Headquarters, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.

3.2.7 *equivalence ~~claims,~~claim, n*—in ~~an~~ equivalence ~~claims,~~claim, two products are claimed to be equivalent in one or more particular feature.

3.2.8 *home use test (HUT), n*—a test where stimuli are evaluated by consumers at home or in the environment typical of the actual use situation; the stimuli preparation and presentation are usually self-administered.
    3.2.8.1 *Discussion*—
Home use test and in-home use test are used interchangeably.

3.2.9 *monadic or single product tests, n*—product tests where only one product is experienced and rated.

3.2.10 *power (statistical), n*—the probability that a sensory claim will be supported, when, in reality, the sensory claim is true.
    3.2.10.1 *Discussion*—
Power = $1 - \beta$, where β is the Type II error rate.

3.2.11 *self-administered questionnaire, n*—questionnaires independently completed by the respondent are referred to as self-administered.

3.2.12 *superiority ~~claims,~~claim, n*—~~a~~ superiority ~~claims assert~~claim asserts a higher level of performance or liking or preference relative to another brand; a superiority ~~claims~~claim can be opposed to a competitive ~~brands~~brand (for example, "cleans better than brand Z") or opposed to an earlier formula of the brand (for example, "now more cleaning power than before"); a superiority claim is supported if a statistically significant proportion of the respondents prefer the advertiser's product.

3.2.13 *unsurpassed ~~claims,~~claim, n*—~~in~~an unsurpassed ~~claims, the~~claim ~~stated~~indicates that the product(s) selected for comparison is not better/higher (or greater than) in some way to the target product(s) for which the analysis is executed.

4. Basis of Claim Classification

4.1 A fundamental step in advertising claim substantiation is creating an explicit statement of the claim prior to actual testing. The statement is then forwarded to all parties concerned in the substantiation process. Concerned parties could include marketing, marketing research, legal, consumer testing, sensory evaluation, research suppliers, etc. The statement is essential as it can encourage collaboration in terms of corporate resources, confirms the selection of appropriate test methods, and has the potential to maximize the chance of making reliable business decisions about the proposed claim, pending the results of substantiation research. Collaboration among all involved parties prior to executing substantiation research is critical in achieving the best results. All involved parties should meet and agree (perhaps several times) prior to implementing the substantiation research.

4.2 Familiarity with the general classification of advertising claims is important in developing clear statements of claims at an early stage and for developing a rational plan for testing. This familiarity also facilitates the process of selecting appropriate testing methods, among the many types of methods available to the consumer/sensory science professional. Each method answers specific questions and may support one type of claim but not another. Therefore, the consumer/sensory science function provides an important source of information and experience in claim substantiation and will provide much of the definition of testing methodology. There are multiple ways to support claims depending on the characteristics of the claim. Two approaches are consumer based and trained panel based evaluations.

4.3 Advertising claims can be divided into two fundamental classifications: comparative and non-comparative. The distinction between the two classifications is whether a comparison is made relative to an existing product (advertiser's or competitor's) or to itself.

4.4 *Comparative Claims*—Comparative claims are designed to compare similarities and differences between two or more products. The basis for comparison can be within the same brand, between two brands, or between a brand and other products in the category.

4.4.1 Comparative claims generally take one of three forms: superiority, equivalence, or unsurpassed. Superiority, equivalence, or unsurpassed claims are further sub-classified into two central areas of application: hedonic and attribute/perception. Hedonics broadly concern measuring the degree of liking and preference; either liking overall or liking that is limited to one or more specific attributes. Attribute/perception claims apply to intensity when measuring one or more specific product attributes.

4.4.2 *Equivalence Claims*—Equivalence claims are claims that assert equivalent levels of performance or liking when comparing a particular product to another product. In general, equivalence claims are made relative to a market/category leader. Two products are not claimed to be exactly equal in one or more particular features; rather, the products are claimed to be sufficiently similar as to be equivalent from a practical perspective.

4.4.2.1 *Hedonic:*
    (1) "Tastes as good as brand X."

4.4.2.2 *Attribute/Perception:*
    (1) "Our product reduces odors as much as brand X."
    (2) "Our product lasts as long as brand X."
    (3) "Our cake is as moist as the leading brand."

4.4.2.3 *Overall Equivalence:*
    (1) "We're just the same, except for the price."
    (2) "You'll never know the difference between us and brand X."

4.4.3 *Unsurpassed Claims*—In unsurpassed claims, the claim stated indicates that the product(s) selected for comparison is not better/higher (or greater than) in some way to the target product(s) for which the analysis is executed. Examples of unsurpassed claims include the following types:

4.4.3.1 *Hedonic:*
    (1) "No other product is better than our product."
    (2) "No other product is more liked for butter flavor."

4.4.3.2 *Attribute/Perception:*
    (1) "No other cake is more moist than ours."
    (2) "No other product has more butter flavor than ours."
    (3) "No other product reduces odors more than our product."
    (4) "No other product lasts longer than our product."
    (5) "No other product is thicker than our product."
    (6) "No other product cleans faster than our product."

4.4.4 *Superiority Claims*—Superiority claims assert a higher level of performance or liking relative to another brand. Superiority claims can be opposed to competitive brands (for example, "cleans better than brand Z") or opposed to an earlier formula of the brand (for example, "now more cleaning power than before"). Examples of superiority claims include:

4.4.4.1 *Hedonic:*
    (1) "Our product tastes better than brand X."
    (2) "Our product tastes better than any other."
    (3) "Our product is preferred over any other brand."

4.4.4.2 *Attribute/Perception:*
    (1) "Our cake is more moist than any other."
    (2) "Reduces odors more than brand X."
    (3) "Lasts longer than any other product."
    (4) "Thicker than brand X."
    (5) "Cleans faster than any other product."

4.4.4.3 In superiority claims, combinations of hedonic claims and attribute/perception claims can sometimes be found, when superiority claims are established based on overall liking and for specific attributes (for example, "Our hosiery is preferred over Brand X for overall liking and it offers more support and comfort.").

4.4.4.4 ~~From a statistical perspective, it can be easier to support claims of unsurpassed than those of superiority or equivalence. Methods for hypothesis-testing will be discussed further in the section on statistical methods.~~

4.5 *Statistical Considerations for Comparative Claim Types:*

5

4.5.1 From a statistical perspective, it may require a smaller consumer sample to support claims of unsurpassed than a claim of superiority or equivalence.

4.5.2 If hypotheses related to multiple attributes are tested independently using the Type I error rate (level α) with the intention of making the claims that are supported, then it elevates the risk of making one or more false claims above the nominal Type I error rate. To control such risk, an appropriate statistical procedure must be used to adjust for multiplicity. Typically, each hypothesis needs to be tested using an adjusted alpha level that is less than the Type I error rate. This testing procedure must be determined and documented prior to data collection. Consult a statistician if required.

4.5.2.1 If hypotheses are related to multiple attributes and no claim will be made unless all of the test results support the claim, then a multiplicity adjustment is not required. This testing procedure must be determined and documented prior to data collection. A longer questionnaire is often needed to test multiple claims. For considerations related to test design, see Section 6 (especially 6.13).

4.5.2.2 If one hypothesis is related to an attribute evaluated by a trained sensory panel and another hypothesis is related to an attribute evaluated by a consumer panel and no claim will be made unless all of the test results support the claim, then a multiplicity adjustment is not required. This testing procedure must be determined and documented prior to data collection.

4.5.3 An advertiser can use a test result to evaluate hypotheses sequentially without adjustment for having conducted multiple tests if the hypotheses are tested using the same approach and differ only in their critical values. Specifically, if one claim is supported, then a subsequent claim can be evaluated without multiplicity adjustment if all possible test results that support the subsequent claim also support the first claim. If using the statistical methods in Section 14, then the following statements hold:

4.5.3.1 Every superiority claim is also an unsurpassed claim. An advertiser can evaluate an unsurpassed claim as a primary hypothesis and a superiority claim as a secondary hypothesis. No claim can be made if the unsurpassed claim is not supported. If the unsurpassed claim is supported, then the unsurpassed claim can be made, regardless of whether the superiority claim is supported. If the superiority claim is supported, then the superiority claim can be made. The unsurpassed test (14.4) and superiority test (14.2) use the same approach and differ only in their critical values. As every test result that supports a superiority claim also supports an unsurpassed claim, it is valid to evaluate both tests using the Type I error rate (level α) without adjusting for multiplicity.

4.5.3.2 Every equivalence claim supported by the test in 14.3 is also an unsurpassed claim that will also be supported by the test in 14.4. These tests use the same critical value for the lower bound. The equivalence test from 14.3 uses an additional critical value for the upper bound. An advertiser can evaluate an unsurpassed claim as a primary hypothesis and an equivalence claim as a secondary hypothesis. No claim can be made if the unsurpassed claim is not supported. If the unsurpassed claim is supported, then the unsurpassed claim can be made, regardless of whether the equivalence claim is supported. If the equivalence claim is supported, then the unsurpassed claim can be made. As every test result that supports an equivalence claim also supports an unsurpassed claim, it is valid to evaluate both the unsurpassed and equivalence claims using the Type I error rate (level α) without adjusting for multiplicity.

4.5.3.3 If using statistical methods other than those recommended in Section 14, then these statements do not necessarily hold, so additional justification is required.

4.5.4 Methods for hypothesis testing can be found in the sections related to the analysis plan (5.2.10) and to statistical methods (Section 14).

4.6 *Non-comparative/Communications Claims*—The objective of the non-comparative/communications claim is to convey something specific about the product, usually a product benefit or difference, and in general, does not seek to provide comparative claims relative to other products. For example, the statement "provides long-lasting flavor" or "smells strong for one month" tells us something about the product, but not in a comparative sense relative to an existing product. These types of claims are common in new product types, but also are used to bring attention to specific product benefits. Examples of non-comparative/communications claims include the following types.

4.6.1 *Hedonic:*
    (1) "Tastes great."
    (2) "Makes your laundry outdoor-fresh."
    (3) "Leaves a long-lasting freshness you will like."

4.6.2 *Attribute/Performance:*
   (1) "Removes odors for 60 days."
   (2) "Leaves glass streak-free."
   (3) "Leaves no residue on surfaces."
   (4) "Works fast."

NOTE 1—In the above attribute examples, some of these could be approached either as a non-comparative claim, since no other product is mentioned, or as a comparative claim versus an appropriate standard (streak-free glass, residue-free surface, odor-free room).

4.7 *Selecting an Appropriate Ad Claims Test*—Product claims made in print or on radio, TV, or the Internet require valid data that supports the intended claim. As with most sensory testing, it is necessary to first identify the project and test objectives for the study. The claim statement should indicate whether the claim is based on consumer or laboratory sensory methods or, in fact, some instrumental or chemical test. Sensory claims for preference or liking ("preferred over the leading brand" or "better than the competition") require consumer tests with the preference or liking questions to support the claim. Claims about product attribute(s) or performance can be based on data from consumers, who are asked about the specific attribute, or from laboratory sensory tests designed to measure the specific attribute(s). In some cases, both types of testing (consumer and laboratory) can be used together to support the same claim. The ad claims team needs to determine the type of claim, the claim statement, the target population, and the aspect(s) of the product that is the focus of the claim. Only then can the test to support the claim generate data with the right focus and weight to support the claim.

5. Consumer Testing: Selection of Study Respondents and Product

5.1 *Study Respondent Sampling:*

5.1.1 Representative sampling of relevant study respondents is a key element of a defensible claim study. The type of claim and the target audience of the advertising are two determining factors for the appropriate respondent sample, within practical considerations of the category and claim. Respondents should be typical users and not be restricted to any specific subset, such as heavy or light users, unless it is a qualified claim about that specific subset. It is also important to distinguish between purchasers or users of the product to ensure sampled respondents align with the claim.

5.1.1.1 Screening based on recent category usage is recommended to identify target respondents. A simple product statement should be used to describe the product to be tested with just enough information so that it is clear what the product is. The product statement should not contain puffery, benefits, aesthetic description, or ingredients, unless these are unusual enough to require an explanation. Sometimes respondents need to be screened to a more narrowly defined criterion. For example, a claim about strawberry yogurt should recruit respondents who like and eat strawberry yogurt, not just any yogurt with fruit. Conversely, respondents typically do not need to be restricted to use of only that product within the category, they may also use alternative products. For the same example, a claim about strawberry yogurt does not mandate that they be exclusive to eating strawberry yogurt. In some cases, the screening criteria needs to be broadened, for example a criterion for inclusion may be that all the participants have the same need, even though that need may be met by products that are designed very differently or for other purposes (for example, widely different products may be used for make-up removal).

5.1.1.2 If a claim is aimed at drawing future category users in addition to present users, future intent to use should be included as a screening criterion. If recent category usage is not applicable (more than the past three to six months), such as with seasonal products or products with long purchase-repeat cycles, or when a product is innovative and part of a newly created category, identifying study respondents based on positive future category usage intent (for example, concept or prototype acceptors) may be necessary. More detail than the simple product statement should only be used when it is necessary to clarify what the product is because of its potential to bias respondents in advance of the claims substantiation test. When the product is new to the market without any comparable product, a formal concept may be used to screen respondents for the claims substantiation testing.

5.1.2 Hedonic or affective (such as liking, acceptance, preference, or purchase intent) claims should always apply to the target user population. Sampling from any population other than the users to whom the claim is focused, such as purchasers, may require a qualified claim to limit its generality. The test protocol should state clearly whether a claim is being made for one or more of the following:

Population	Examples
Product recommenders	"doctors prefer..."
Purchasers	Pet food by pet owners
	"Choosy moms choose Jif"[A] is a claim specific to the purchaser, the user is the child
The ultimate consumer of a product	"now more refreshing"

[A] Jif is a registered trademark of the J.M. Smucker Company.

5.1.3 Perception claims are based on product attribute(s) or performance, or both, that are detectable in the target user population. Claims that state or imply detectability (for example, 'more chocolate flavor'), include study participants from the target user population. Note, trained and validated panel data can be used, if data related to the claim from the target population already exists, and it can be demonstrated that the trained panel's measurements can be used to predict the target population (see 10.3.2 and 10.3.3).

5.1.4 For category usage claims, respondents may be recruited by screening for brand usage, but care should be taken during screening to ensure respondents cannot determine which brands are targeted for testing. This can be accomplished by providing a large list of brands with the brand or brands of interest embedded in the questionnaire. Brand usage and frequency of use data also can be collected to help match the incidence of these variables in the respondent sample and target population. Target users can be defined by their responses to several questions, including:

5.1.4.1 "What one brand of this product type do you use most often?"

5.1.4.2 "What brands have you used in the last (insert time period appropriate for the category)?"

5.1.4.3 "What brand have you used most often in the last 'x' months?"

5.1.4.4 If the frequency of use is one of the criteria for recruitment, then the respondent also may be asked how often they use the product or how many times they have purchased the product within a specific time frame. More questionnaire discussion can be found in 6.9.

5.1.5 Potential allergies or sensitivities to product ingredients should be assessed based on the product being tested and nature of exposure. Participants in product studies need to be screened by self-report for potential allergens, sensitizers, or lifestyle factors, or combinations thereof, that may exclude them from the test (for example, tree nuts, wheat, or special diets for religious or ethical reasons for food and beverage products, nickel, and some perfume components for non-food products).

5.2 *Study Respondent Sampling Techniques:*

5.2.1 The type of claim should be kept in mind when determining ~~sample size.~~ the number of respondents. For example, ~~an~~ equivalence ~~claims~~claim may require more respondents than a superiority ~~claims~~claim (see Section 14~~), and some objective claims (for example, "this product has more...") can be substantiated through descriptive analysis by a trained panel (see Section)~~. Furthermore, the number of respondents required to represent the range of relevant respondent characteristics may be larger than the minimum that is required to achieve desired ~~10).~~statistical power.

5.2.2 The ~~demographics~~characteristics of the respondent sample should be consistent with those of the target population (i.e, about whom the claim is being made). The ~~demographics~~characteristics may include the population in terms of age, gender, product usage, culture, and language. ~~A constrained demographic sample, such as a single gender~~ For example, a constrained demographic sample should be employed when it is consistent with the stated claim and normal product usage (for example, a particular gender or age).

5.2.3 Using quotas on relevant respondent characteristics is important to achieve a match between sampled respondents and the intended target population. ~~Demographic~~ Relevant respondent information ~~must~~should be collected to demonstrate the appropriateness of the sample ~~and respondent~~ of the target population. Respondents chosen should reflect the target population. For example, representation of age and gender should reflect the target population.

5.2.4 Recruiting of the respondent sample must be stated in the test protocol and should be as objective as possible. Develop and document a rationale for both the number of respondents and for the inclusion and exclusion criteria. If the ~~sample size~~ number of respondents is less or more than determined by statistical criteria alone, this should be included in the rationale. Consumer product testing respondents are usually obtained from stratified (not convenience) sampling procedures. Quotas are often used because they tend to produce a ~~consumer~~ sample of respondents that is fit for the purpose of testing with the target user population

of interest. When there are multiple testing locations (see 5.2.7.2), then these quotas are applied within each of the locations. Probability sampling for consumer product testing is rarely practical because it would consider all consumers in the population of interest as potential participants in the test. In reality, many of these consumers cannot be located or are unwilling to participate.

5.2.4.1 Screening criteria should not be revealed to potential respondents during recruitment, and the standard security screening questions (for example, whether family members work in advertising or marketing or other related fields, including that of the test product) should be included.

5.2.5 *Sources for Selecting Respondents*—Potential respondents may be obtained from companies that provide testing services or sell marketing information. These databases are built using a variety of techniques, without targeting a particular manufacturer's consumers. In some cases, a company may maintain its own branded database of product users. If using these databases, it is important to consider potential biases, for example, the repeated use of the same individuals, or the influence of the known sponsor of the research.

5.2.6 Respondents should be screened specifically for the claims substantiation test to ensure they currently meet the inclusion and exclusion criteria in the study design. To minimize over testing over-testing with the same consumers,respondents, it is common to screen for past study participation, excluding those who have tested a specific product in the past three months, or tested in the product category in the past six months. Other approaches to minimize sampling bias may include conducting testing across a range of days of the week and times of the day, and varying the locations where potential respondents are recruited.

5.2.7 *Geographic Area Recruiting*—Some claims will require geographic considerations as a function of the target population and the nature of the claim. The study design and rationale may need to include variables that reflect the geography of the claim. Variation in product use may be due to geographic variability in demographics. Product performance may be impacted by environmental factors such as temperature, humidity, water hardness, pollution, or hours of sunshine. For example, use of certain sauces and spices may vary geographically, laundry detergent performance and preferences may be impacted by water hardness, hair product performance may be impacted by humidity, and suntan lotion use may vary by geographic region.

5.2.7.1 Preference Some preference claims and usage claims may have a potential for geographic and demographic dependencies. These claims may vary For example, such a dependency exists if consumer preference or usage varies by region or by socioeconomic factors, such as urban versus suburban versus rural. The evidence for or against such dependencies could come from patterns in product sales, or usage, or both. These considerations need to be included in the respondent sampling rationale when they apply.

5.2.7.2 When geographic region is assumed to be a factor relevant to a claim, the geography of respondents should be consistent with the scope of the claim. A rationale should be developed when testing country-wide or regionally, ensuring a fair representative sample of the defined target population in multiple markets dispersed across the geography stated or implied in the claim. For example, some studies conducted by means of Central Location Testing (CLT, Various 6.6) in the USA sample from the four major regions (West, East, Central, and South), with two cities per region. An example for Home Use Testing (HUT, 6.6), is using zip codes or postal codes, to ensure testing in multiple markets across the geography. Other approaches for selection of test locations are possible; they need to be included in the rationale. Some examples of how test sites could be selected follow.
 (1) Recruiting Example Based on Geography: National Claim—An advertiser intends to make a national claim for a product in the United States. A consumer test is conducted via Central Location Testing (6.6.1) in four major regions (West, East, Central, and South) with two cities per region, with a sufficiently large number of consumers included from each of the cities. In general, it is best to recruit a large enough sample of consumers in each city to have the range of relevant characteristics represented in the consumer sample and to achieve the desired statistical power (see Section 14). If the consumer sampling requirement is met by having 40 consumers per city, this would yield a total sample of 320 consumers (40 consumers × 8 cities). If the consumer sampling requirement is met by having 60 consumers per city, this would yield a total sample of 480 consumers (60 consumers × 8 cities). Consumer sampling depends on both the range of relevant characteristics needed to be included in the consumer sample to achieve representativeness and the risk of challenge.
 (2) Recruiting Example Based on Geography: National Claim—An advertiser intends to make a national claim for a product in the United States. A consumer home use test (HUT; 6.6.2) is conducted using zip codes to ensure that a sufficiently large number of consumers are included in multiple markets across geographies. In general, it is best to recruit a large enough sample of consumers to have the range of relevant characteristics represented in the consumer sample and to achieve adequate statistical power (see Section 14). To achieve the targeted number of consumers, it will likely be necessary to recruit some additional consumers. Consumer sampling depends on both the range of relevant characteristics needed to be used in the recruiting/screening of the consumer sample to achieve representativeness and the risk of challenge. Relevant characteristics refer to the criteria used in selecting respondents to serve in a sensory/consumer product test. For example, these criteria can pertain to factors such as demographic characteristics, product usage ("users of automatic dish detergents," "users of bottled barbeque sauce"), purchasing

habits or potential to purchase a product or product category, flavor likers, attitudes, or behaviors relevant to the need for a product (for example, frequent gym goers desiring gym apparel that reduces odors; parents with young children who need healthy snacks in portable packaging).

(3) Recruiting Example Based on Geography: Regional Claim—An advertiser intends to make a regional claim in the western United States for a product that is distributed only in the western United States. Three cities in western U.S. states where the advertiser and competitor products have comparable market share are selected. The consumers sampled are users who are representative of consumers with respect to relevant factors (i.e., demographic, socio-economic and product usage). A test is conducted via CLT with a sufficiently large number of consumers from each of these three cities. For example, a comparative taste claim pertaining to a mayonnaise product that is distributed only west of the Mississippi River would require mayonnaise consumers to be sampled from cities in states west of the Mississippi River.

(4) Recruiting Example Based on Brand Development Index in Limited Geographies—An advertiser intends to make a regional claim for a hand cream product that is designed to protect and soothe hands exposed to cold temperatures. Sales are concentrated in regions in northern U.S. states, mostly in northern Minnesota, upstate New York and northern Maine. Since there are three regions where the Category Development Index (CDI) is highest, a test is conducted via consumer CLT in three cities in these regions, one city per region. The cities are selected so that taken together each brand's Brand Development Index (BDI) is approximately proportional to that brand's BDI for the regions in which the claim will be made. The test includes a sufficiently large number of consumers from each city to represent target users of this category of products.

5.2.8 *Recruiting Example Based on Market Share*—If market share is a criterion, be cautious when selecting markets and ensure that the test adequately represents the people residing in the geographic territory on which the claim is based. In categories with strong geographic differences in market share, the total market share should be approximated by representing high, low, and average share markets in the study. ~~Regional sample sizes~~ The number of respondents per region may vary, reflecting their contribution in terms of population density, but not heaviness of usage.

5.2.9 *Recruiting Examples Based on Product Usage*—Respondent selection may be based on product usage or product category usage rather than geographical considerations. Examples follow.

5.2.9.1   Sensory testing is needed to support a claim that a product delivers relief from itchy eyes, the primary need for selecting a representative sample of respondents is to have users of the category independent of geography. A large sample of users of products in this category can serve as a representative consumer sample in this case.

5.2.9.2 A product claim pertains to hair coloring products. Only users of product in the hair coloring category need to be included in the respondent sample.

5.2.10 *Analysis Plan*—Claims substantiation is confirmatory testing, not exploratory testing. Once a target population is defined in the study protocol and represented adequately by sampling, results from the total respondent sample are used for substantiating a claim. Results among some subgroup may not correspond to overall results because ~~sample size~~ the number of respondents in subgroups are smaller, and therefore may not be as statistically reliable. Moreover, since there is a risk of false positives and false negatives in testing any hypothesis, analysis of multiple subgroups will increase the overall error rate. Therefore, the analysis plan reflects the intended claim(s) and any subgroup testing should be determined before testing begins (for example, ensure ~~adequate sample size~~ an adequate number of respondents for any analysis and minimize the number of hypotheses to test). Any subgroup analyses that were not planned and properly designed may not be a sound analytical practice for claims substantiation (see Section 14). If the number of respondents who complete the test is lower than what was planned, researchers need to provide a rationale for the shortage or a rationale for why the achieved number of respondents is sufficient to provide adequate data to support the claim.

5.2.11 Be prepared to defend ~~your~~ the documented rationale for respondent sample criteria if challenged.

5.3 *Selection of Products:*

5.3.1 If a test is being conducted to support a competitive claim that is not brand-specific (for example, versus "other leading brands"), then the competitive brands should be the two brands with the highest national market share. If the market is highly fractionated, such that the top two national brands control less than 50 % of the market, then more competitors must be included in the test. Either the three leading national brands or any brand that is among the top two in the four major geographic regions of the country must be tested. Unless the product is tested against brands representing at least 85 % of the national market, it is recommended that claims should be made against specific brands in lieu of general superlative claims. Eighty-five percent (85 %) of the market is defined as all products within said category, including the brand making the claim.

5.3.2 Competitive brands should be in the same market segment as the brand for which the claim is being made. If a brand straddles market segments, then products most similar in a reasonable competitive context should be used.

5.3.3 When competing products are sold in more than one form, the products being tested must be of the same form or in the form most relevant to the claim. If a powdered drink mix is being compared with a competitor's product that also comes in a powdered drink mix and as a reconstituted liquid, both brands would have to be tested in their reconstituted from powdered forms. The specific directions for preparation given on each product must be followed. If there is substantial crossover use of different forms, a claim involving different forms may be desired. The forms tested must be stated explicitly as part of the claim, for example, "instant tastes as good as ready-made."

5.4 *Sampling of Products When Both Products are Currently on the Market:*

5.4.1 For central location consumer tests, commercial products to be used for competitive claims testing should be purchased at the end of the distribution chain to ensure the product is representative of the product the consumer would purchase. Some products are made at different or multiple manufacturing sites. In those instances, the product should be purchased from a distribution center that services the particular test areas.

5.4.2 For other test methods in which the test product is manufactured at one location, samples can be purchased from any high volume store. Products should be sourced at the same time from the same store(s) in each local testing area. Products should reflect the choice available to local consumers. Care should be taken to include a variety of production sites and dates that typically are found on the retail shelf.

5.4.3 In cases where competitive products are not sold in the same stores (for example, fast food restaurants and private label products) test products should be sourced as close in time as possible from locations that reflect choices available to local consumers. It is important that the geographic identity of samples match that of local test participants. This way, if national products manufactured in more than one site have been formulated differently to appeal to regional differences in sensory preferences, appropriate products will be tested against relevant regional competitors. It is critical that all information regarding product sourcing be documented.

5.4.4 Competitive products should be purchased in the standard size package with the highest unit volume or in similar size, or both, to the test product. Trial size and club-store oversized product packages should not be used unless the package meets the specific target of the claim.

5.4.5 Every effort should be made to obtain competitive products of representative freshness found in the marketplace. All products in the test should be of typical age. A freshly-made product should not be compared against a product nearing its expiration date.

5.5 *Handling of Products When Both Products are Currently on the Market:*

5.5.1 After procurement but prior to testing, handling, length of storage, and storage conditions of all products must be identical and consistent with normal consumer practice.

5.5.2 Competitive samples must not show any signs of mishandling or abuse. If products become non-homogeneous during handling, in that they cannot be returned to their original state (precipitates may be returned to solution, but fractured pieces cannot be made whole), then test samples should be remedied for such defects. For example, the last serving or two from a box of cereal that may have a disproportionate share of fines should be discarded or screened.

5.5.3 To minimize the likelihood of product recognition by respondents, manufacturers sometimes try to "blind" the competitive product. Manipulations beyond labeling the original package should be approached with extreme caution. Repackaging of product would need to be supported by instrumental and sensory tests demonstrating no impact on the product. Any alteration of the product itself to minimize recognition could potentially impact acceptability and should be applied with the utmost discretion. It may be feasible to remove a product from its identifying package, but altering the structure of a product, such as grinding cereals to mask their shape, may change a product beyond the point where the competitive assessment is credible. When a product is instantly recognizable by its appearance, shape, or design, then cognitive factors due to brand recognition or previous experience with the product may contribute to the ratings obtained in the study.

5.6 *Sampling of Product Not Yet on the Market:*

5.6.1 If the manufacturer's product is not yet on the market at the time of testing, the product should represent commercial production, and either be typical retail age of competitive products or expected age of the product when the cycle of the manufacturer's distribution is observed. The competitive product should be selected to represent average retail age at the time of testing. If a suitable product is not available in the test city, the product should be sourced from a nearby location.

5.6.2 To ensure that the claimed benefit of the new product results from the product itself and not from special handling during limited scale production, it is desirable, but may not always be practical, for the new product to have been made at the production facility. A new product, therefore, should be made at its intended manufacturing site, preferably on the same equipment and under normal operating conditions that will be used to manufacture the product. If pilot plant material must be used for claim support, then supplemental testing, for example, discrimination test for similarity, must be conducted to demonstrate that the claim benefits extend to material made at the production facility.

5.7 *Sample Preparation/Test Protocol:*

5.7.1 To minimize bias, it is essential that all samples for testing are prepared and served in a manner that will have limited impact on the perception of the products and in a manner that treats all of the products fairly.

5.7.2 For claims substantiation tests in particular, samples should be prepared and served under reasonably realistic conditions, i.e., in a manner consistent with normal consumer practice. Samples should not be prepared in any fashion that would mask or alter various product characteristics.

5.7.3 All samples should be tested blind and with unbiased codes, such as three-digit codes. The respondents should have no leading or biasing information about the products that they are testing or about the overall objective of the study.

5.7.4 A decision must be made regarding the manner in which the samples will be presented to the respondents. For example, the samples can be served as pairs or one at a time (monadic presentation). Differences among samples are more likely to be detected when two or more samples are presented together; however, monadic presentation generally is considered more representative of the consumer experience.

5.7.5 The order of sample presentation must also be considered prior to testing and this must be designated according to a statistical design. Various psychological factors can influence judgment, for example, the impact for which the following order effects must be accounted:

5.7.5.1 *Context/Contrast Effect*—As defined in Terminology E253.

5.7.5.2 *Positional Bias*—Respondents may be more sensitive to differences in specific samples in a series, such as the first or last sample.

5.7.5.3 *Pattern Effect*—Any pattern in order will be detected quickly.

5.7.5.4 *Ceiling Effects*—This typically occurs when the majority of the scores occur towards the top of a rating scale. When the products are well-liked, there is not a sufficient amount of scale available to the respondents to differentiate the products. Variation in rating scores is compressed, making mean-based statistical tests misleading. Therefore, analysis should be performed using a more robust statistical model that does not have distributional requirements and is less prone to outlier influence such as multinomial logistic regression.

5.7.6 It is essential to balance the order of presentation to distribute these effects across all products.

5.7.7 The test and questionnaire should be designed to be free of all forms of bias. Bias during testing may come from the samples, the test protocol, including the questionnaire, or the test environment, or a combination thereof. Other sections of this guide discuss these issues.

6. Test Design—Consumer Testing

6.1 Monadic designs are those in which a single product is rated by respondents at a time.

12

6.1.1 Sequential monadic designs require each respondent to evaluate products one at a time and in consecutive order.

6.1.2 Protomonadic tests consist of providing one product, obtaining ratings of that product on a variety of attributes, removing the first product, and replacing with a second product. No monadic ratings are obtained on the second product; instead, a paired-comparison test is conducted.

6.2 Comparative test designs are those in which two or more products are presented to the same respondents to compare the products to each other.

6.3 Comparative claims imply, but are not limited to, comparative designs, where each respondent evaluates two or more products. For comparative claims, paired comparisons are used most frequently. Simultaneous presentation provides the most direct comparison of the products. In some situations, sequential presentation may be needed that introduces execution and sensitivity issues, so there should be a rationale for choosing a sequential (monadic) presentation.

6.3.1 In cases where there are multiple products to be compared, the respondents may be able to evaluate all of the products (complete balanced block design) or a subset of products (an incomplete block design) or only a single product (monadic design). When the products are evaluated in subsets, overlapping product blocks may be constructed using techniques such as Balanced Incomplete Blocks (BIBS) and Partially Balanced Incomplete Blocks (PBIBS). These Incomplete Block designs may require specialized analysis procedures to construct the correct averages, as outlined in Cochran and Cox (1)[4] and other statistical references.

6.4 Since monadic testing is not the most direct method for making comparisons, it is not always the most desirable approach. Nevertheless, sometimes it may be the only practical method to support comparative claims. For example, some products may require long periods of repeated usage to provide a consumer benefit, which can undermine the ability to make direct comparisons. In this case, product performance can be assessed by giving each product to a different group of consumers and conducting statistical analysis on the ratings. In monadic designs, respondents, as well as products, contribute to the total variation, rendering it less sensitive and larger differences or larger sample sizes are required for significance. It is critical that the groups be matched adequately.

6.5 Non-comparative claims may be supportable by either monadic or sequential-monadic test designs. While a monadic rating may provide a measure free from influences inherent in multi-product, sequential-monadic designs, either approach is sufficient to meet the "reasonable basis" required to make a claim.

6.5.1 Qualitative research, such as focus groups, is not acceptable for claims support since one cannot project their findings to a larger population of consumers.

6.5.2 Both central location (CLT) and home use (HUT) test methods can be acceptable, depending on the specifics of the category and usage. CLTs include all locations other than respondents' homes. These locations may include sensory facilities, mall facilities, field sites, supplier's premises, community centers, or others. Each type of location has some benefits and limitations that must be taken into consideration when projecting results.

6.6 *Data Collection Strategies:*

6.6.1 *Central Location Testing (CLT)*—This method of testing provides maximum control over product preparation and usage. Central location testing assures that the participant actually evaluated the product in question and provides his or her own opinion immediately following evaluation, rather than relying on past usage or recollection. Blind testing often precludes the need to repackage product. In addition, CLTs can provide direct product comparisons, isolate specific attributes, such as color or crunchiness, vanilla flavor, and so forth, and accommodate complex evaluation protocols. They are appropriate for superiority, equivalence, or unsurpassed claims.

6.6.1.1 Key limitations are that central location tests usually involve a single product exposure with small amounts of product under conditions that may not closely duplicate typical usage. Questions about whether such exposure can exaggerate trivial differences or whether CLTs provide a basis for forming a preference have been raised. Other limitations that can be controlled

---

[4] The boldface numbers in parentheses refer to the list of references at the end of this standard.